Title

Gut bacteria responding to dietary change encode sialidases that exhibit preference for red meat-associated carbohydrates.

Journal

Nature microbiology, 4(12)

ISSN

2058-5276

Authors

Zaramela, Livia S
Martino, Cameron
Alisson-Silva, Frederico
et al.

Publication Date

2019-12-01

DOI

10.1038/s41564-019-0564-9

Peer reviewed

HHS Public Access

Author manuscript

Nat Microbiol. Author manuscript; available in PMC 2020 March 23.

Published in final edited form as:

Nat Microbiol. 2019 December ; 4(12): 2082–2089. doi:10.1038/s41564-019-0564-9.

Gut bacteria responding to dietary change encode sialidases that exhibit preference for red meat-associated carbohydrates

Livia S. Zaramela[1,#], **Cameron Martino**[1,2,#], **Frederico Alisson-Silva**[3,4,#,§], **Steven D. Rees**[5], **Sandra L. Diaz**[3,4], **Léa Chuzel**[6], **Mehul B. Ganatra**[6], **Christopher H. Taron**[6], **Patrick Secrest**[3,4], **Cristal Zuñiga**[1], **Jianbo Huang**[5], **Dionicio Siegel**[5], **Geoffrey Chang**[5], **Ajit Varki**[3,4], **Karsten Zengler**[1,7,8,*]

[1]Department of Pediatrics, University of California, San Diego

[2]Bioinformatics and Systems Biology Program, University of California, San Diego

[3]Department of Medicine and Cellular & Molecular Medicine, University of California, San Diego

[4]Glycobiology Research and Training Center (GRTC)

[5]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego

[6]New England Biolabs

[7]Departmemt of Bioengineering, University of California, San Diego

[8]Center for Microbiome Innovation, University of California, San Diego.

Abstract

Dietary habits have been associated with alterations of the human gut resident microorganisms contributing to obesity, diabetes, and cancer[1]. In Western diets, red meat is one of the frequently eaten foods[2], but long-term consumption has been associated with increased risk of disease[3,4]. Red meat is enriched in *N*-glycolylneuraminic acid (Neu5Gc) that cannot be synthesized by humans[5]. However, consumption can cause Neu5Gc incorporation into cell surface glycans[6], especially in carcinomas[4,7]. As a consequence, an inflammatory response is triggered when Neu5Gc-containing glycans encounter circulating anti-Neu5Gc antibodies[8,9]. Although bacteria can utilize free sialic-acids as a nutrient source[10-12], it is currently unknown if gut microorganisms contribute to releasing Neu5Gc from food. We found that a Neu5Gc-rich diet induces changes in the gut microbiota, with *Bacteroidales* and *Clostridiales* responding for the most. Genome assembling of mouse and human shotgun metagenomic sequencing identified bacterial sialidases with previously

*Corresponding author.

Author Contributions: LSZ, CM, FAS, AV, and KZ conceptualized the study. LSZ, FAS, and KZ wrote the manuscript with input from all authors. LSZ and CM performed and analyzed the microbiome experiments. LSZ and SLD performed the enzymatic characterization. LSZ, FAS, and PS performed the animal work. SDR performed the protein expression and crystallization experiments assisted by JH and DS. CZ performed the metabolic model analysis. LC, MBG, and CT constructed the fosmid library and performed the compost sialidase activity assays. CT, GC, AV, and KZ provided resources and supervised the study.

#These authors contributed equally to this work.

§Current affiliation: Paulo de Goes Institute of Microbiology, Federal University of Rio de Janeiro, Brazil.

Declaration of Interests: L.S.Z., C.M., F.A.-S., S.R., S.L.D., G.C., A.V. and K.Z. have filed a patent application (number pending) that claims the use of sialidases to reduce Neu5Gc.

unobserved substrate preference for Neu5Gc-containing glycans. X-ray crystallography revealed key amino-acids potentially contributing to substrate preference. Additionally, we verified that mouse and human sialidases were able to release Neu5Gc from red meat. The release of Neu5Gc from red meat using bacterial sialidases could reduce the risk of inflammatory diseases associated with red meat consumption, including colorectal cancer[4] and atherosclerosis[13].

Red meat is enriched in Neu5Gc that cannot be synthesized by humans due to an evolutionary loss of a functional CMP-Neu5Ac hydroxylase (CMAH)[5]. Metabolic incorporation of Neu5Gc into tissues requires glycosidically bound-Neu5Gc for reasons currently unknown. In contrast, free-Neu5Gc is utilized by gut microbes or cleared rapidly by the kidneys through the urine[14] (Supplementary Fig. 1). Intestinal bacteria can release host-derived sialic-acids from mucosal mucins and glycolipids by expressing sialidases[15-18]. It is also known that gut commensal and pathogenic bacteria can utilize sialic-acids as carbon source[10-12]. However, to our knowledge, every bacterial sialidase tested prefer Neu5Ac to Neu5Gc[19-21]. How bound-Neu5Gc is metabolized in the gut is therefore currently unknown. Once free, sialic-acids can be taken up through membrane-associated transporters and utilized as carbon, nitrogen, or energy sources, or used to sialilate bacterial cell surface glycans[16]. In addition, changes in the intestinal concentration of sialic-acids, for example induced by inflammation, can alter the expression of bacterial genes involved in sialic-acid catabolism promoting intestinal dysbiosis[22]. Due to the importance of sialic-acids in microbe-host interaction within the gut, we investigated if a Neu5Gc-rich diet could provoke changes in bacterial metabolism and alter the gut microbiome. We compared the microbiota composition in fecal samples of *Cmah*$^{-/-}$[23] that were fed a sialic-acid free (soy) diet, a Neu5Gc-rich porcine submaxillary mucin (PSM) diet, or a Neu5Ac-rich edible bird's nest (EBN) diet[14] (Supplementary Table 1).

Fecal contents were scraped from the mouse colons and the microbiome of *Cmah*$^{-/-}$ mice and WT mice fed soy, PSM, or EBN diets was determined through 16S rRNA gene amplicon sequencing. Bray-Curtis dissimilarity showed a significant difference in the bacterial genotypes present in *Cmah*$^{-/-}$ and WT mice, indicating that a mouse's inability to synthesize endogenous Neu5Gc significantly impacted microbiome composition (Fig. 1a) and substantiated the need for a human-like *Cmah*$^{-/-}$ mouse model in our experimentation. We found that changes in the microbial composition in *Cmah*$^{-/-}$ mice were diet-dependent, with *Clostridiales* and *Bacteroidales* contributing significantly to the variations observed amongst the diets (Fig. 1b; Supplementary Fig. 2). Independent of genotype, the microbiome of PSM-fed mice was significantly less diverse compared to the microbiome of those fed soy and EBN diets (p-*value* < 0.05) (Supplementary Fig. 3a). Human-like *Cmah*$^{-/-}$ mice revealed similar taxonomic profiles at the family level amongst the three diet groups (Supplementary Fig. 3b). However, at the genus level, *Helicobacter, Intestinimonas*, and Candidatus *Saccharibacteria genera incertae sedis* were significantly enriched in the EBN group compared to PSM (Figure 1c). Moreover, *Bacteroides*, *Barnesiella*, *Clostridium* group III, *Parabacteroides, Roseburia*, and *Turicibacter* were significantly enriched in the PSM group compared to EBN (p-*value* < 0.05) (Fig. 1c, Supplementary Fig. 4). *Bacteroides* were previously associated to efficiently metabolize carbohydrates from plant- as well as animal-based food due to their diverse enzymatic repertoire[24,25]. Additionally, by computationally

simulating 773 metabolic models from human gut microbiome members[26], we found that member of the *Bacteroidetes*, including *B. fragilis, B. cacae, B. thetaiotaomicron*, were one of the most efficient microorganisms utilizing sialic-acids as a carbon source (Supplementary Fig. 5, Supplementary Table 2, and Supplementary Discussion). Shotgun metagenomic DNA sequencing was performed to evaluate the enzymatic repertoire for utilizing carbohydrates in the microbial community from *Cmah*$^{-/-}$ mice fed soy, PSM, or EBN diets, and raw reads were aligned to a carbohydrate active enzymes (CAZymes) dbCAN-seq database[27]. A Principal Coordinate Analysis revealed a near significant clustering in gene function between diet types (ANOSIM R=0.246, *p-value*=0.06), with a higher similarity between the CAZymes on soy and EBN diets (Supplementary Fig. 6). Additionally, the examination of individual sialidase genes revealed several diet-dependent sialidases (Fig. 1d). To evaluate the sialidases genes present in the microbiome, the combined metagenomes were co-assembled and 51 genome bins containing 21 sialidase genes were identified (Supplementary Fig. 7, Supplementary Table 3). Amongst the bins with annotated sialidase genes, bin13, whose closest relative was *Bacteroides thetaiotaomicron* (Supplementary Table 3), was the most abundant in PSM compared to EBN diets (Fig. 1e). Bin13 contains five sialidases (sialidase23, sialidase24, sialidase26, sialidase60, and sialidase65). Sialidase26 exhibited high amino acid sequence conservation (81% identity) to sialidase CUA18247.1 from *B. fragilis*, the most abundant sialidase protein in the PSM diet (Fig. 1d, Supplementary Fig. 8).

The increased abundance of bin13 in PSM suggests its prominent role in Neu5Gc metabolism in the gut. As such, bin13 sialidases might possess biochemical properties that make them better suited for Neu5Gc release. To test this notion, each of the five sialidases genes in bin13 were PCR amplified (Supplementary Fig. 9) and heterologously expressed *in vitro*, purified, and assayed for its substrate preference. Sialidase activity measurements were performed using different enzyme concentrations (0.5-10 μg) at three different pHs (6.5, 7.0, and 8) (Fig. 2a, Supplementary Fig. 10). Four out of five sialidases showed preferential Neu5Gc activity in at least one of the pHs tested (Fig. 2a). To the best of our knowledge, no previously characterized exo-sialidases have been shown to prefer Neu5Gc over Neu5Ac[21,28-30] (Supplementary Discussion). Additionally, we tested *in vivo* sialidase activity in fresh fecal samples. Clarified fecal pellets from mice fed with PSM tend to preferentially release Neu5Gc compared to clarified fecal pellets from mice fed with Soy (Fig. 2d).

The most compelling enzyme from the substrate specificity study in all tested conditions was sialidase26. Sialidase26 possessed protein sequence motifs characteristic of the GH33 family of sialidases (Supplementary Fig. 11, Supplementary Fig. 12). Despite this substrate preference, sequence residues that are predicted to interact with terminal sialic-acids in the catalytic site are highly conserved with structurally studied sialidases exhibiting unknown Neu5Gc preference over Neu5Ac (Supplementary Fig. 13, Supplementary Discussion). To elucidate the structural underpinnings of Neu5Gc preference, we used X-ray crystallography to determine the structure of sialidase26 both alone (PDB 6MRX, 2.0Å resolution) and in complex with the inhibitors DANA-Ac (*N*-acetyl-2,3-dehydro-2-deoxyneuraminic acid) and DANA-Gc (*N*-glycolyl-2,3-dehydro-2-deoxyneuraminic acid) (PDB 6MRV, 1.8Å resolution, and PDB 6MYV, 2.2Å resolution, respectively) (Supplementary Table 4). Sialidase26

structure is common to GH33 sialidases (Fig. 2b), including Y509 nucleophilic engagement of C5 following E398 charge activation, D228 acid-base catalysis of the glycosidic bond at C5, and C2 stabilization by an Arg triad (R203, R414, R478). However, typical sialic-acid stabilizing interactions are lost, including Glu engagement of the glycerol moiety C7-C9 (T397 in sialidase26), C10 stability by an Arg residue (now a Leu), and inward movement of W507 into the binding pocket (though this is restored in the DANA-Gc co-crystal structure). Co-crystallization of sialidase26 with DANA-Gc indicated an overall fold and ligand placement similar to that of DANA-Ac (Fig. 2c). The hydroxyl-group at the end of the C5 acetamido-group of DANA-Gc is pointed towards the binding pocket residues, forming H contacts (hydrogen bonds) with D271 and likely increasing its stability. Amino acid substitutions at this position to Leu (D271L) or Asn (D271N) significantly lowered sialidase protein activity and eliminated the Neu5Gc preferential cleavage (Fig. 2a). This, in combination with the changes described above, likely explains sialidase26's preference for Neu5Gc.

To extend our findings to humans, we reanalyzed fecal shotgun metagenomes from the Hadza (Supplementary Table 5), a genetically distinct indigenous ethnic group that resides in remote Tanzania[31]. The Hadza are hunter-gatherers and change their diet periodically throughout the year according to food availability. In dry seasons, their diet is enriched in meat and tubers, whereas in wet seasons it consists largely of honey and berries[31,32]. Raw reads from Hadza metagenomes were mapped to bin13. We observed that bin13 was significantly more abundant in microbiome samples taken during the dry season compared to the wet season (Fig. 3a). To evaluate the sialidases genes present in the Hadza microbiome, the combined metagenomes were co-assembled and 24 genome bins containing 51 sialidase genes were identified (Supplementary Fig. 14, Supplementary Table 6). The binHz19, whose closest relative was *Alistipes* sp., was equally abundant in microbiomes from both seasons (Fig. 3b, Supplementary Table 6). BinHz19 also contained a sialidase (sialidaseHz136) with the greatest sequence similarity to sialidase26 identified in our mouse study, suggesting widespread distribution of this sialidase homolog amongst mammals (Supplementary Fig. 15). SialidaseHz136 showed preferential activity for Neu5Gc over Neu5Ac in all tested conditions (Fig. 3c). X-ray crystallography of SialidaseHz136 (PDB 6MNJ, 2.2Å resolution) exhibited similar predicted engagement of conserved residues with sialic acid substrate when compared to sialidase26 (Fig. 3d), despite a shift in the acetamido-interacting Asp residue. This difference supports the notion that sialidaseHz136 is capable of metabolizing structurally diverse glycans, a concept that would be of benefit to an individual with seasonal variations in diet. To additionally determine if sialidase26 and sialidaseHz136 can release Neu5Gc from food sources directly, we test their activity on beef, pork, and PSM chow as substrate. Both enzymes showed pronounced sialidase activity by releasing Neu5Gc from all food tested (Fig. 3e).

Our findings indicate that sialidases exhibiting Neu5Gc preference can be widespread in the mammalian intestine. To determine if Neu5Gc-specific sialidases are restricted to enteric bacteria, we utilized functional metagenomic screening to identify exo-sialidases with a Neu5Gc specificity in a terrestrial environment. A fosmid library containing ~40kb inserts of total environmental DNA isolated from soil from an organic composting facility was constructed in *E. coli*. Lysates from 5,376 clones were screened for hydrolysis of 4MU-α-

Neu5Gc (Supplementary Fig. 16). Clones showing activity were tested subsequently and re-tested for their ability to cleave 4MU-α-Neu5Ac. Two clones (C19 and C22) showed significant hydrolysis of 4MU-α-Neu5Gc but only minor activity on 4MU-α-Neu5Ac (Fig. 3f). The DNA sequence of cloned inserts from C19 and C22 each encoded a single bacterial sialidase gene. Both enzymes showed exo-sialidase activity and had a marked preference for Neu5Gc hydrolysis *in vitro* compared to several other known bacterial exo-sialidases (Fig. 3f). Analysis of the deduced amino acid sequences of C19 and C22 showed they are highly similar to each other (56% identity), both belong to the GH33 family of bacterial exo-sialidases (Supplementary Fig. 17), and are most similar to proteins from terrestrial bacteria of the genera *Rhodopirellula* and *Verrucomicrobia*. Considered together, these data support the conclusion that exo-sialidases with Neu5Gc preference also exist in terrestrial environments, where decomposing biological material from animal sources might require such preference.

In summary, we identified several sialidases with previously undescribed preference for Neu5Gc, which are enriched in the gut microbiota of mice and humans upon consumption of a Neu5Gc-rich diet. Previous studies from our group showed that dietary free Neu5Gc does not get incorporated in *Cmah*$^{-/-}$[14] mice. Thus, the cleavage of Neu5Gc from foods entering the gut can potentially prevent incorporation of this non-human sugar into the colon tissue. We also hypothesize that the gut microbiome with an underrepresentation of bacteria with Neu5Gc-preferring sialidases could result in increased xenosialitis and be one potential contributing factor to inflammation-mediated promotion of diseases. While further *in vivo* characterization needs to be done, our results lay the foundation to define a strategy for translation of pre- or probiotics to prevent incorporation or to eliminate Neu5Gc from tissues of red meat eaters therefore reducing the risk of xenosialitis and other diseases associated with red meat consumption.

Methods

In vivo sampling, feeding, and animal diet

Wild type (WT) lineage C57BL/6 was purchased from Harlan Laboratories and human-like *Cmah*$^{-/-}$ mice generated as previously described[23]. In brief, the generation of the human-like *Cmah*$^{-/-}$ transgenic mice was performed by target deletion of exon 6 of the *cmah* gene (similar deletion that evolutionary occurred in humans) using loxP sites and Cre recombinase expression in embryonic stem cells. Transgenic mice were generated by the University of California, San Diego (UCSD), Transgenic Mouse Core. All animal experiments were approved by the UC San Diego Institutional Animal Care and Use Committee (IACUC) under the protocol number S01227. All animals were maintained in the University of California, San Diego vivarium according to Institutional Animal Care and Use Committee (IACUC) guidelines, with 12 hours diurnal lighting and access to food and water *ad libitum*. Sample processing and analysis were not blinded at any step. Samples size was chosen based on the cost, mice, diet and vivarium availability, and minimum number of samples to perform statistical analysis. WT and *Cmah*$^{-/-}$ mice were raised in the same vivarium room and fed with the same water source. The cages were kept side by side in the same cage rack to minimize external influence on the gut microbiome. Age and sex matched

female mice used in the study were maintained in sialic acid free soy based diet (Dyets, Inc.; 110951) from weaning until 8 to 10 weeks of age to prevent previous exposure to sialic acid. To evaluate the effect of the dietary Neu5Gc in the gut microbiome, sex matched *Cmah*$^{-/-}$ mice with 10 weeks of age were caged in three groups of 5 mice each and fed during 4 weeks with the same soy based diet either enriched in Neu5Gc (porcine submaxillary mucin - PSM)[14], Neu5Ac (edible bird nest - EBN)[5], or kept in sialic acid free-soy only as control. The animals were euthanized in CO_2 chamber and the colon tissues were cut open with blunted scissors for fresh collection of fecal samples by scraping it straight from the tissue. Colonic fecal scrapings from five mice of each diet type were used for 16S rRNA gene amplicon and 3 mice of each diet type were used for metagenomic shotgun sequencing analysis. The feeding protocol was chosen based on previous evidence from our group showing that feeding *Cmah*$^{-/-}$ mice with PSM over a period of weeks can cause mouse tissue incorporation of Neu5Gc at levels histologically similar to the levels seen in adult humans who have eaten red meat for many years[14]. The PSM diet was prepared to contain 250 μg of Neu5Gc per gram of chow to mimic the amount of Neu5Gc present in beef, the most consumed form of red meat in the western diet.

Monosaccharide and Amino acid composition of the diets

Monosaccharide composition of PSM and EBN diets were analysed by High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection (HPAEC-PAD) using Dionex ICS-3000 system (ThermoFisher Scientific) equipped with CarboPac PA1 column 4 mm × 250 mm, 4μm, with a 4 mm × 50 mm Guard. Briefly, 500 μg of each diet were dissolved in 200 μL of Milli-Q water. The samples were hydrolysed by adding an equal volume of 4N Trifluoroacetic acid (final concentration of 2N TFA) at 100°C for 4 hours. The hydrolysed samples were centrifuged at 400*g* for 2 min and evaporated under a flow of dry nitrogen. Once dried, samples were resuspended in 200 μl of Milli-Q water and 50% of each sample was injected. The separation of monosaccharide peaks was achieved by using the following solvents and gradient conditions:

Solvents	Gradient setting			
A: Water	Time	%A	%B	%C
B: 100 mM NaOH with 5 mM NaOAc	0	84	16	0
C: 100 mM NaOH with 250 mM NaOAc	20	84	16	0
	21	0	100	0
Pulsed Amperometric Detector	31	0	100	0
Waveform: Standard Quad	32	84	16	0
	50	84	16	0

The amino acid composition of both diets was performed by GC-MS tBDMS (dimethyl-tert-butylsilyl) derivatives quantitation as previously described[33]. Both HPAEC-PAD and GC-MS analysis were performed by the Glycotechnology Core at the Glycobiology Research and Training Center – UCSD. https://medschool.ucsd.edu/research/GRTC/services/glycoanalytics/Pages/default.aspx.

16S rRNA sequencing and analysis

Total genomic DNA were extracted using MoBio PowerFecal DNA isolation kit (MoBio, Carlsbad, CA, USA) following the manufacturer's instructions. Purified DNA was amplified and processed according to the Illumina 16S protocol (https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). 16S rRNA libraries were generated from 3 or 5 biologic replicates and 3 independent experiments per diet group. Libraries were quality assessed using quantitative PCR (qPCR) and Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and subsequently sequenced using two MiSeq 600 cycle kits (Illumina). Adapters were trimmed from the Illumina data using Trimmomatic version 0.36[34]. 16S analysis was performed with usearch denoising and the RDP (Ribosomal Database Project) 16S rRNA database version 16. Sequences were analysed using the Usearch v10[35] following the MiSeq 2×250 16S V4 pipeline (https://www.drive5.com/usearch/manual/pipe_examples.html). In brief, paired-end reads were merged using fastq_mergepairs (-fastq_maxdiffs 10; -fastq_pctid 10). Sequences with a distance-based similarity of 97% or greater were grouped into OTUs using cluster_otus (-minsize 2). OTU table were rarefied to 10,000 observations per sample. OTU-based microbial diversity and dissimilarity metrics were estimated using R package vegan v2.5-2. Bray-Curtis distance were used to beta-diversity analysis. Vector fitting with each of the PCoA ordinations was performed using the function *envfit* from R package vegan v2.5-2. Significant vectors were selected following the criteria (R >= 0.7 and p-values =< 0.01). Bacteria genus with relative abundance below than 1% were not considered for differential abundance analysis. Statistical differences in abundance between diets were calculated using the nonparametric Wilcoxon rank sum test with Holm correction for multiple hypotheses when appropriated.

Shotgun metagenome sequencing and analysis

As described above, total genomic DNA were extracted using MoBio PowerFecal DNA isolation kit (MoBio, Carlsbad, CA, USA) following the manufacturer's instructions. Purified DNA from three biological replicates per diet group was prepared for shotgun metagenomic sequencing using the Nextera XT library preparation method with the average fragment size of 450 bp (Illumina, San Diego, CA, USA). Libraries were quality assessed using quantitative PCR (qPCR) and a Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and subsequently sequenced using MiSeq 2×250 bp cycle kits (Illumina). In average, 2.5 million non-mouse reads were generated per library. Adapters were trimmed from the Illumina data using Trimmomatic v0.36[34]. Samples were filtered of possible human and mouse contamination by aligning the trimmed reads against reference databases using Bowtie2 v2-2.2.3[36] with the following parameters (-D 20 -R 3 -N 1 -L 20 –very-sensitive-local). Overlapped reads were merged using Flash version 1.2.11[37]. Merged and unmerged reads were assembled using Spades v3.12.0 with the following parameters (-k 21,33,55,77,99,127 --meta --merge)[38]. Differential binning was performed using MetaBat2 v2.12.1, with minimum contig length of 1500 bp[39]. Bin quality (completeness and contamination) was evaluated using CheckM v1.0.7[40]. Taxonomic classification (closest phylogenetic neighbour) was assessed using RAST online tool[41]. In brief, RAST (Rapid Annotation using Subsystem Technology) uses a set of unique proteins to assign the closest related neighbour. Genome annotations were performed using Prokka v1.11 with default

parameters. Amino acid sequences of all genes identified using Prokka were aligned to sialidase sequences obtained from Uniprot until the present date (October, 2018). The alignment was performed using Diamond v0.8.24 with the following parameters (blastx -k 5 -f 6 --evalue 0.001). Relative bin abundance was obtained by dividing the total number of reads aligned per bin by the total number of reads aligned in all bins. For this analysis, non-mouse trimmed reads were aligned to the binned genomes using Bowtie2 v2-2.2.3[36] with the parameters set by the flag --very-sensitive. Bins with relevant genes were compared for relative abundance between sample groups using the Wilcoxon rank sum test. Carbohydrate active enzymes (CAZymes) database[27] were used to identify diet-dependent sialidases. Non-mouse trimmed reads were aligned against sialidase amino acid sequences using Diamond v0.8.24 using the follow parameters: -k 5 -f 6 --evalue 0.001. When indicated, counts per million were used to normalize the number reads aligned by protein sequence. The total number of reads by protein sequence were rarefied to 5,000 observations per sample using the R function *rrarefy* from R package vegan v2.5-2. Bray-Curtis distance were used to CAZyme diversity analysis. Amino acid sequence similarities were evaluated using BLAST (Basic Local Alignment Search Tool) online tool and pairwise sequence alignments were generated using Clustal Omega online tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). Sialidases per bacterial genome were obtained from PATRIC (Pathosystems Resource Integration Center) database using command-line interface *P3-scripts*.

Analysis of previously published shotgun metagenomic data

Previously published stool shotgun metagenomic data from Hadza hunter-gatherer individuals[31] were obtained from Sequence Read Archive (SRA) repository under the project IDs PRJNA392012 and PRJNA392180. 40 shotgun metagenomic data from individuals were analysed, in which 20 samples were collected during the wet season and 20 during the dry season (Supplementary Table 5). All published shotgun metagenomic data were processed as described in the previous section.

Protein Expression and Assay

Target sialidase sequences from shotgun metagenomic data were PCR amplified from genomic DNA isolated as described above or ordered from Integrated DNA Technologies (IDT), subcloned into a pET19b expression vector with a C-terminal 10×His tag and N-terminal truncation to remove any signal peptide sequence (predicted by SignalP 4.1, CBS), and transformed into BL21(DE3) *E. coli* (MilliporeSigma) using established heat-shock methods. Cells were grown to OD 0.6-0.8 (OD 600) in multiple 1 L cultures at 37 °C, and induced overnight at 25 °C with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Harvested cells were resuspended in lysis buffer (50 mM HEPES pH 8.0, 50 mM NaCl, and 1 mM TCEP (tris(2-carboxyethyl)phosphine hydrochloride))) with DNaseI and hen egg white lysozyme, lysed with a TS-Series cell disruptor (Constant Systems, Inc.) at 15 KPSI (Kilo-Pound per Square Inch), and spun for 45 minutes at 186,000×g with a Ti45 ultracentrifugation rotor (Beckman Coulter, Inc.) to remove cell debris. Purification was performed as below and based on purification of a putative *Bacteroides* neuraminidase as provided by the Protein Structure Initiative (BACCAC_01090, Joint Center for Structural Genomics, to be published), with modifications to imidazole stringency based on the sialidase purified. Supernatant was loaded on a 5-mL HisTrap Ni affinity column (nickel-

charged columns for high resolution histidine-tagged protein purification) on an Akta Explorer purification system (GE Healthcare Life Sciences) with 20-40 mM imidazole added, washed with Running Buffer (50 mM HEPES pH 8.0, 300 mM NaCl, 40-60 mM imidazole, 10% glycerol, and 1 mM TCEP), and eluted with Elution Buffer (20 mM HEPES pH 8.0, 300 mM imidazole, 10% glycerol, and 1 mM TCEP). Samples were concentrated using 10-30 kDa Amicon centrifugal filters (MilliporeSigma) at 1500×g to 1 mL, and desalted over a 5-mL. Desalting column using the Akta system into Desalting Buffer (20 mM HEPES pH 8.0, 200 mM NaCl). Resulting protein sample was diluted as needed for functional studies.

Assay for sialidase activity

***In vitro* activity:** Sialidases purified as described above were quantified using SDS-PAGE image analysis with BSA (Bovine Serum Albumin) references (Bio-Rad) and absorbance using the Nanophotometer P330 (Implen), with extinction coefficients calculated using ExPASy Translate (https://web.expasy.org/translate/). Sialidase activity assays were performed in a dilution series; 0.5μg, 2.5μg, 5μg, and 10μg of each enzyme were incubated with equal amount of human-like *Cmah*$^{-/-}$ mouse serum (Neu5Ac: 1428.64 pmoles/μL, Neu5Gc: 0 pmoles/μL) and WT mouse serum (Neu5Ac: 196.29 pmoles/μL, Neu5Gc: 1337.45 pmoles/μL) for 1 hour at 37 °C. An additional 10 μg of each enzyme was inactivated by heat for 5 minutes at 95 °C. The samples were kept at −20 °C until derivatization and analysed by HPLC as described below. The DMB reagent was made with the following recipe: 14 mM DMB (1,2-diamino-4,5-methylenedioxybenzene, Sigma D4787), 18 mM sodium hydrosulfite (Sigma 157953), 1.0 M 2-mercaptoethanol (Sigma M3148), and 40 mM trifluoroacetic acid (Sigma T6508), and it was incubated at 50 °C for 2.5 h. The DMB-derivatized samples were analyzed on a Dionex Ultra3000 HPLC System using a Phenomenex Gemini 5μ C18 250 × 4.6-mm HPLC column at room temperature, eluted in isocratic mode with 85% water, 7% methanol, 8% acetonitrile. The same protocol was utilized to evaluate the sialidases activity in three different pH ranges.

***In vivo* activity using clarified fecal samples:** Fecal pellets were collected from *Cmah*$^{-/-}$ mice with 10 weeks of age caged in two groups of 4 mice and 3 mice each, fed during 2 weeks with PSM[14] or sialic acid free soy, respectively. All the animals' information and maintenance are the same as described in the section "*In vivo* sampling". Fresh fecal pellets were homogenised (10% w/v) in Tris-HCL pH 7 buffer with protease inhibitor (Roche 11836170001) using Pellet Pestle Cordless Motor (Kimble 749540). Samples were spined down, 4.000g for 3 minutes and the supernatant was transferred to SpinX filters (Costar 8160). Samples were centrifuged 14.000g for 10 minutes and the flow through was transfered to a clean tube. Total protein concentration was determined using Pierce BCA (Bicinchoninic Acid) Protein Assay (Thermo Scientific 23225). To measure sialidase activity in clarified fecal samples, 2μg of total protein per sample was incubated with equal amount of human-like *Cmah*$^{-/-}$ mouse serum and WT mouse serum for two hours at 37 °C. An additional 2μg of total protein per sample was inactivated by heat for 5 minutes at 95 °C. Samples were cleaned up using 10 kDa Amicon centrifugal filters (MilliporeSigma), 14.000×g for 25 min at 4 °C. Samples were lyophilized and resuspended in 50 μL of sterile

water. Derivatization and HPLC analysis was performed as described in the section "*In vitro* activity".

Food sources: Beef (New York Steak) and pork (breakfast sausage) were purchased at Whole Foods Market and cut in small pieces using sterile blades. PSM chow was powdered using crucible and pistil. Beef, pork, and PSM were homogenised (10% w/v) in Tris-HCL pH 7 buffer using Pellet Pestle Cordless Motor (Kimble 749540). 50 μL of each suspension was incubated with 2μg of either sialidase26 or sialidaseHz136 for two hours at 37 °C. An additional 2μg of each sialidase was inactivated by heat for 5 minutes at 95 °C. Reactions were performed in triplicates. Samples were cleaned using 10 kDa Amicon centrifugal filters (MilliporeSigma), 14.000×g for 25 min at 4 °C. Derivatization and HPLC analysis was performed as described in the section "*In vitro* activity".

Neu5Gc2en Synthesis

Neu5Gc2en (DANA-Gc) was synthesized as previously published[42-45] with minor modifications. Briefly, Neu5Gc (Sigma-Aldrich) was treated with Dowex 50W-X8 (H+) resin in MeOH for 20 hours at 20 °C to form the methyl ester. This ester was then treated with acetic anhydride and pyridine for 42 h at 20 °C to generate the peracetylated methyl ester, which was purified by column chromatography (50:l $CHCl_3$- MeOH). This sample was treated with TMSOTf under dry nitrogen at 0 °C in MeCN for 6 h to induce elimination. The unsaturated compound was purified by chromatography (toluene/acetone, 3:1 → 2:1). The acetyl groups were cleaved by treatment with NaOH in MeOH over 12 hours followed by neutralization with H+ resin.

Crystallization and Structure Determination of Sialidase26 and SialidaseHz136

Purified sialidase26 was concentrated to 8 g/L and set on sitting drop trays in 1:1 volume ratios with mother liquor (20% PEG (PolyEthylene Glycol) 6000, 0.1 M Tris-HCl pH 8.0) at 16 °C. For ligand co-crystals, concentrated sialidase26 was incubated for one hour at room temperature with 5 mM N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid (Neu5Ac2en or DANA, Sigma-Aldrich) or DANA-Gc (synthesized). Crystals appeared after 3-4 days and grew to full size in 8-10 days. Crystals were soaked briefly in mother liquor supplemented with 10% glycerol and flash-frozen with liquid nitrogen. Purified sialidaseHz136 was crystallized similarly (20% PEG 3350, 0.1 M Bis-Tris-Propane pH 7.5, 0.2 M sodium citrate). X-ray diffraction data were collected at 100 K at the Lawrence Berkeley National Laboratory Advanced Light Source (8.2.1 and 8.2.2) at a single wavelength. Preliminary diffraction data were collected at the Stanford Synchrotron Radiation Lightsource and Advanced Photon Source. All diffraction data were indexed and integrated with XDS (X-ray Detector Software) (2018-Jun-08) or MOSFLM (Measurement of Oscillation FiLMs), processed with AIMLESS v0.7.2, and truncated with CTRUNCATE within the CCP4 v7.0.073 suite of programs [46-49]. Phases were estimated via molecular replacement in PHENIX.PHASER v1.13, using a previously published model of an uncharacterized Bacteroides-derived sialidase with high sequence homology to sialidase26 (PDB 4q6k) as a search model. Models underwent rigid-body and restrained positional refinement using PHENIX.REFINE in the PHENIX software suite v1.13[50] against a maximum likelihood target function, alternated with manual inspection against electron density maps in Coot

v0.8.9.1[51]. Geometry restraints for DANA were generated using PHENIX.eLBOW[52], with manual inspections in Coot and refined in the final rounds of refinement, which also included the application of hydrogens to their riding positions and simulated annealing. For the co-crystal structure of sialidase26 and DANA-Gc, diffraction data from two crystals with similar unit cell parameters were processed together in AIMLESS and multi-crystal averaging performed with PHENIX.MULTI_CRYSTAL_AVERAGE for five cycles following Phaser of each before refinement proceeded with PHENIX.REFINE. The resulting refinement statistics for each model are included in table S4. Figures displaying crystal packing were prepared using PyMOL v1.8.3.2 (http://www.pymol.org), and atomic coordinates and structure factors were deposited with the Protein Data Bank (accession codes,6MNJ, 6MRV, 6MRX, and 6MYV). Ramachandran statistics for Sialidase26 alone (6MRX), with DANA (6MRV), or with DANA-Gc (6MYV), and SiaHz136 (6MNJ), are, respectively, outliers (0.39, 0.19, 0.24, and 0.39%) and favored (95.67, 96.06, 95.87, 95.27%). ALS beamline v.8.2.2 were used for all datasets save 6MYV, which was collected on 8.2.1. Wavelengths used for the same datasets are, respectively, 0.99999, 0.99995, 1.00003, and 0.99994Å. All were collected at 100.0 K. No Cβ deviations were observed for any dataset.

Enzyme Discovery - Metagenomic DNA library construction

Environmental DNA was isolated from soil obtained from a commercial organic composting facility in Hamilton, MA (Brick End Farms) by phenol:chloroform extraction and isopropanol precipitation. A fosmid library was produced using the CopyControl™ Fosmid Library Production Kit (Lucigen Corporation, Middleton, WI) as recommended. Briefly, DNA was end-repaired and size-selected using a 1% Low Melting Point agarose gel run overnight at 35 V. DNA fragments from 30-70 kb were isolated from the gel using 1 U of β-agarose I (New England Biolabs, Ipswich, MA) for each 100 μL of melted agarose. The end-repaired and size-selected DNA was ligated to the pCC1-FOS cloning vector. Resulting clones were packaged in phage particles. *Escherichia coli* EPI300 T1R cells were transfected with the packaging reaction and plated on LB agar medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g MgCl2•6H2O, chloramphenicol 12.5 μg/mL, 2 mL 2M NaOH and 20 g of agar per litter) and incubated overnight at 37°C. A total of 5376 colonies were archived in fourteen 384-well plates in sterile 20% (v/v) glycerol.

Enzyme Discovery - Screening for sialidase activity

The compost metagenomic library (theoretically encoding ~215,000 environmental genes) was differentially screened with fluorogenic 2′-(4-methylumbelliferyl)-α-D-N-glycolylneuraminic acid (4MU-α-Neu5Gc) (Sussex Research, Ottawa, CA) and 2′-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid 4MU-α-Neu5Ac (Toronto Research Chemicals, Toronto, CA) substrates. In a primary screen, library clones were grown in 384-well plates containing 50 μL LB liquid cultures (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g MgCl2-6H2O, 2 mL of 2M NaOH per liter, containing chloramphenicol 12.5 μg/mL and 1× inducing solution (Lucigen Corporation)) overnight at 37 °C. Fifty microliters of Y-per lysis buffer (Thermo Fischer Scientific, Waltham, MA) containing 40 μg/mL of 4MU-α-Neu5Gc was added to each well. The mixtures were incubated overnight at 37 °C in a static incubator. Fluorescence at λex = 365 nm and λem = 445 nm was read

with a SpectraMax Plus 384 Microplate Reader (Molecular Devices, Sunnyvale, CA) at 6 h, 24 h and 48 h. Positive clones were defined as those showing fluorescence greater than 3 standard deviations above the mean background, and each was re-archived in a fresh 384-well plate in sterile 20% (v/v) glycerol. Each of the positive clones was grown and comparatively re-screened in separate assays containing 4MU-α-Neu5Ac and 4MU-α-Neu5Gc substrates (reactions run in duplicate). Two clones designated C19 and C22 showed significant activity on 4MU-α-Neu5Gc but only minor activity on 4MU-α-Neu5Ac and were subjected to further study.

Reconstructed metabolic models analysis

Genome-scale network reconstructions combine detailed biochemical and physiological information, providing insights into the metabolism for subsequent manipulation strategies[53] or to control metabolism[54,55]. The scope of these models encompasses the characterization of the metabolic behaviour of target microorganisms. We evaluated growth phenotypes of microorganisms of the gut microbiota associated with sialidase metabolism (EC: 3.2.1.18). Seventeen microorganisms of the gut microbiota containing sialidases were identified by scanning the repository of the gut microbiome metabolic models[26]. Growth rates were simulated using flux balance analysis (FBA). All metabolic models were constrained using a western diet (45% fat, 35% carbohydrate, 20% protein), containing experimental constraints for 20 sugars, 24 fiber-related metabolites, 12 fatty acids, 20 amino acids, and 88 minerals, vitamins and other metabolite. Experimental constraints are reported in great detail in the gut microbiota repository[26]. All metabolic models were simulated using the Gurobi Optimizer Version 5.6.3 (Gurobi Optimization Inc., Houston, Texas) solver in MATLAB (The MathWorks Inc., Natick, MA) with the COBRA (COnstraint-Based Reconstruction and Analysis) Toolbox[56]. Additionally, the contribution to growth of the metabolites associated with sialidase activity (e.g. N-acetylneuraminate) was determined using shadow prices simulations[56].

Statistics and Reproducibility

Statistical analysis was performed using R programming language. The statistical significance of differential relative abundance (16S rRNA amplicon and shotgun metagenomics) was computed using no-parametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypotheses. The statistical significance of sialidase activity was computed using student's 2-tailed t test. The significance levels are indicated as follow: $p\text{-}value \leq 0.05$ (*), $p\text{-}value \leq 0.01$ (**), and $p\text{-}value \leq 0.001$ (***). All analysis was performed in biologically independent animals or independent experiments as indicated in the text. Samples were not randomized; all data were used in the analysis.

Data Availability

Sequencing data supporting the findings of this study are available under accession number PRJNA505660. X-ray crystallographic data that support the findings of this study have been deposited in the RCSB (Research Collaboratory for Structural Bioinformatics) Protein Data Bank (accession codes: 6MRX, 6MRV, 6MYV, and 6MNJ).

Code Availability

Custom scripts to generate figure and for statistical analysis and any additional information are available from the corresponding author upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank all Zengler-, Varki-, and Chang-lab members for helpful discussion. Research was supported in part by the National Institutes of Health under award number R01GM32373 (to AV) and by the National Science Foundation under award number IOS-1444435 (to GC). CM was supported by grants from the National Institutes of Health, USA (NIH grant T32GM8806) and by a Chancellor's Research Excellence Scholarship (UCSD). FAS was partly supported by the Program Science Without Borders Bex 9254-13-7 - Capes Brazil.

References

1. Hall AB, Tolonen AC & Xavier RJ Human genetic variation and the gut microbiome in disease. Nat. Rev. Genet 18, 690–699 (2017). [PubMed: 28824167]
2. Mann N Dietary lean red meat and human evolution. Eur. J. Nutr 39, 71–79 (2000). [PubMed: 10918988]
3. Etemadi A et al. Mortality from different causes associated with meat, heme iron, nitrates, and nitrites in the NIH-AARP Diet and Health Study: Population based cohort study. BMJ 357, 1–11 (2017).
4. Alisson-Silva F, Kawanishi K & Varki A Human risk of diseases associated with red meat intake: Analysis of current theories and proposed role for metabolic incorporation of a non-human sialic acid. Mol. Aspects Med 51, 16–30 (2016). [PubMed: 27421909]
5. Samraj AN et al. A red meat-derived glycan promotes inflammation and cancer progression. Proc. Natl. Acad. Sci 112, 542–547 (2015). [PubMed: 25548184]
6. Tangvoranuntakul P et al. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. Proc. Natl. Acad. Sci 100, 12045–12050 (2003). [PubMed: 14523234]
7. Samraj AN, Laubli H, Varki N & Varki A Involvement of a non-human sialic acid in human cancer. Front. Oncol 4, 1–13 (2014). [PubMed: 24478982]
8. Varki A Uniquely human evolution of sialic acid genetics and biology. Proc. Natl. Acad. Sci 107, 8939–8946 (2010). [PubMed: 20445087]
9. Dhar C, Sasmal A & Varki A From "serum sickness" to "xenosialitis": Past, present, and future significance of the non-human sialic acid Neu5Gc. Front. Immunol 10, 807 (2019). [PubMed: 31057542]
10. Almagro-Moreno S & Boyd EF Sialic acid catabolism confers a competitive advantage to pathogenic *Vibrio cholerae* in the mouse intestine. Infect. Immun 77, 3807–3816 (2009). [PubMed: 19564383]
11. McDonald ND, Lubin J-B, Chowdhury N & Boyd EF Host-derived sialic acids are an important nutrient source required for optimal bacterial fitness *in vivo*. MBio 7, e02237–15 (2016). [PubMed: 27073099]
12. Lewis AL & Lewis WG Host sialoglycans and bacterial sialidases: A mucosal perspective. Cell. Microbiol 14, 1174–1182 (2012). [PubMed: 22519819]
13. Kawanishi K et al. Human species-specific loss of CMP-*N*-acetylneuraminic acid hydroxylase enhances atherosclerosis via intrinsic and extrinsic mechanisms. Proc. Natl. Acad. Sci 201902902 (2019). doi:10.1073/pnas.1902902116
14. Banda K, Gregg CJ, Chow R, Varki NM & Varki A Metabolism of vertebrate amino sugars with N-glycolyl groups: Mechanisms underlying gastrointestinal incorporation of the non-human sialic

acid xeno-autoantigen *N*-glycolylneuraminic acid. J. Biol. Chem 287, 28852–28864 (2012). [PubMed: 22692204]

15. Almagro-Moreno S & Boyd EF Insights into the evolution of sialic acid catabolism among bacteria. BMC Evol. Biol 9, 118 (2009). [PubMed: 19470179]
16. Almagro-Moreno S & Boyd EF Bacterial catabolism of nonulosonic (sialic) acid and fitness in the gut. Gut Microbes 1, 45–50 (2010). [PubMed: 21327116]
17. Li J & McClane BA NanI sialidase can support the growth and survival of *Clostridium perfringens* strain F4969 in the presence of sialyated host macromolecules (mucin) or Caco-2 cells. Infect. Immun 86, e00547–17 (2018). [PubMed: 29203541]
18. Tailford LE et al. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. Nat. Commun 6, 7624 (2015). [PubMed: 26154892]
19. Kim S, Oh DB, Kang HA & Kwon O Features and applications of bacterial sialidases. Appl. Microbiol. Biotechnol 91, 1–15 (2011). [PubMed: 21544654]
20. Juge N, Tailford L & Owen CD Sialidases from gut bacteria: a mini-review. Biochem. Soc. Trans 44, 166–175 (2016). [PubMed: 26862202]
21. Chokhawala HA, Yu H & Chen X High-throughput substrate specificity studies of sialidases by using chemoenzymatically synthesized sialoside libraries. ChemBioChem 8, 194–201 (2007). [PubMed: 17195254]
22. Huang YL, Chassard C, Hausmann M, Von Itzstein M & Hennet T Sialic acid catabolism drives intestinal inflammation and microbial dysbiosis in mice. Nat. Commun 6, 8141 (2015). [PubMed: 26303108]
23. Hedlund M et al. *N*-glycolylneuraminic acid deficiency in mice: implications for human biology and evolution. Mol. Cell. Biol 27, 4340–4346 (2007). [PubMed: 17420276]
24. David LA et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559–563 (2014). [PubMed: 24336217]
25. Flint HJ, Scott KP, Duncan SH, Louis P & Forano E Microbial degradation of complex carbohydrates in the gut. Gut Microbes 3, 289–306 (2012). [PubMed: 22572875]
26. Magnúsdóttir S et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat. Biotechnol 35, 81–89 (2017). [PubMed: 27893703]
27. Huang L et al. DbCAN-seq: A database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res. 46, D516–D521 (2018). [PubMed: 30053267]
28. Owen CD et al. Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont *Ruminococcus gnavus*. Nat. Commun 8, 2196 (2017). [PubMed: 29259165]
29. Inoue S et al. A unique sialidase that cleaves the Neu5Gcα2 → 5-OglycolylNeu5Gc linkage: Comparison of its specificity with that of three microbial sialidases toward four sialic acid dimers. Biochem. Biophys. Res. Commun 280, 104–109 (2001). [PubMed: 11162485]
30. Davies LRL et al. Metabolism of vertebrate amino sugars with *N*-glycolyl groups: Resistance of α2-8-linked *N*-glycolylneuraminic acid to enzymatic cleavage. J. Biol. Chem 287, 28917–28931 (2012). [PubMed: 22692207]
31. Smits SA et al. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science (80-.). 357, 802–805 (2017).
32. Schnorr SL et al. Gut microbiome of the Hadza hunter-gatherers. Nat. Commun 5, 3654 (2014). [PubMed: 24736369]
33. Wood PL, Khan MA & Moskal JR Neurochemical analysis of amino acids, polyamines and carboxylic acids: GC-MS quantitation of tBDMS derivatives using ammonia positive chemical ionization. J. Chromatogr. B Anal. Technol. Biomed. Life Sci 831, 313–319 (2006).
34. Bolger AM, Lohse M & Usadel B Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]
35. Edgar RC Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461 (2010). [PubMed: 20709691]
36. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012). [PubMed: 22388286]

37. Magoč T & Salzberg SL FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963 (2011). [PubMed: 21903629]
38. Nurk S, Meleshko D, Korobeynikov A & Pevzner PA MetaSPAdes: A new versatile metagenomic assembler. Genome Res. 27, 824–834 (2017). [PubMed: 28298430]
39. Kang DD, Froula J, Egan R & Wang Z MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165 (2015). [PubMed: 26336640]
40. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P & Tyson GW CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055 (2015). [PubMed: 25977477]
41. Aziz RK et al. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics 9, 75 (2008). [PubMed: 18261238]
42. Li Y et al. Identifying selective inhibitors against the human cytosolic sialidase NEU2 by substrate specificity studies. Mol. Biosyst 7, 1060–1072 (2011). [PubMed: 21206954]
43. Bill Cai T, Lu D, Landerholm M & Wang PG Sialated diazeniumdiolate: A new sialidase-activated nitric oxide donor. Org. Lett 6, 4203–4205 (2004). [PubMed: 15524443]
44. Ercégovic T & Magnusson G Highly stereoselective α-sialylation. Synthesis of GM3-saccharide and a bis-sialic acid unit. J. Org. Chem 60, 3378–3384 (1995).
45. Numata M, Sugimoto M, Shibayama S & Ogawa T A total synthesis of hematoside, α-NeuGc-(2→3)-β-Gal-(1→4)-β-Glc-(1→1)-Cer. Carbohydr. Res 174, 73–85 (1988). [PubMed: 3378233]
46. Battye TGG, Kontogiannis L, Johnson O, Powell HR & Leslie AGW iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. Sect. D Biol. Crystallogr 67, 271–281 (2011). [PubMed: 21460445]
47. Evans P Scaling and assessment of data quality. Acta Crystallogr. Sect. D Biol. Crystallogr 62, 72–82 (2006). [PubMed: 16369096]
48. Winn MD et al. Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr 67, 235–242 (2011). [PubMed: 21460441]
49. Kabsch W XDS. Acta Crystallogr. Sect. D Biol. Crystallogr 66, 125–132 (2010). [PubMed: 20124692]
50. Adams PD et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D Biol. Crystallogr 66, 213–221 (2010). [PubMed: 20124702]
51. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr 66, 486–501 (2010). [PubMed: 20383002]
52. Moriarty NW, Grosse-Kunstleve RW & Adams PD electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. Acta Crystallogr. Sect. D Biol. Crystallogr 65, 1074–1080 (2009). [PubMed: 19770504]
53. Tan J, Zuniga C & Zengler K Unraveling interactions in microbial communities - from co-cultures to microbiomes. J. Microbiol 53, 295–305 (2015). [PubMed: 25935300]
54. Zengler K & Zaramela LS The social network of microorganisms - How auxotrophies shape complex communities. Nat. Rev. Microbiol 16, 383–390 (2018). [PubMed: 29599459]
55. Zuñiga C, Zaramela L & Zengler K Elucidation of complexity and prediction of interactions in microbial communities. Microb. Biotechnol 10, 1500–1522 (2017). [PubMed: 28925555]
56. Schellenberger J et al. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. Nat. Protoc 6, 1290–1307 (2011). [PubMed: 21886097]

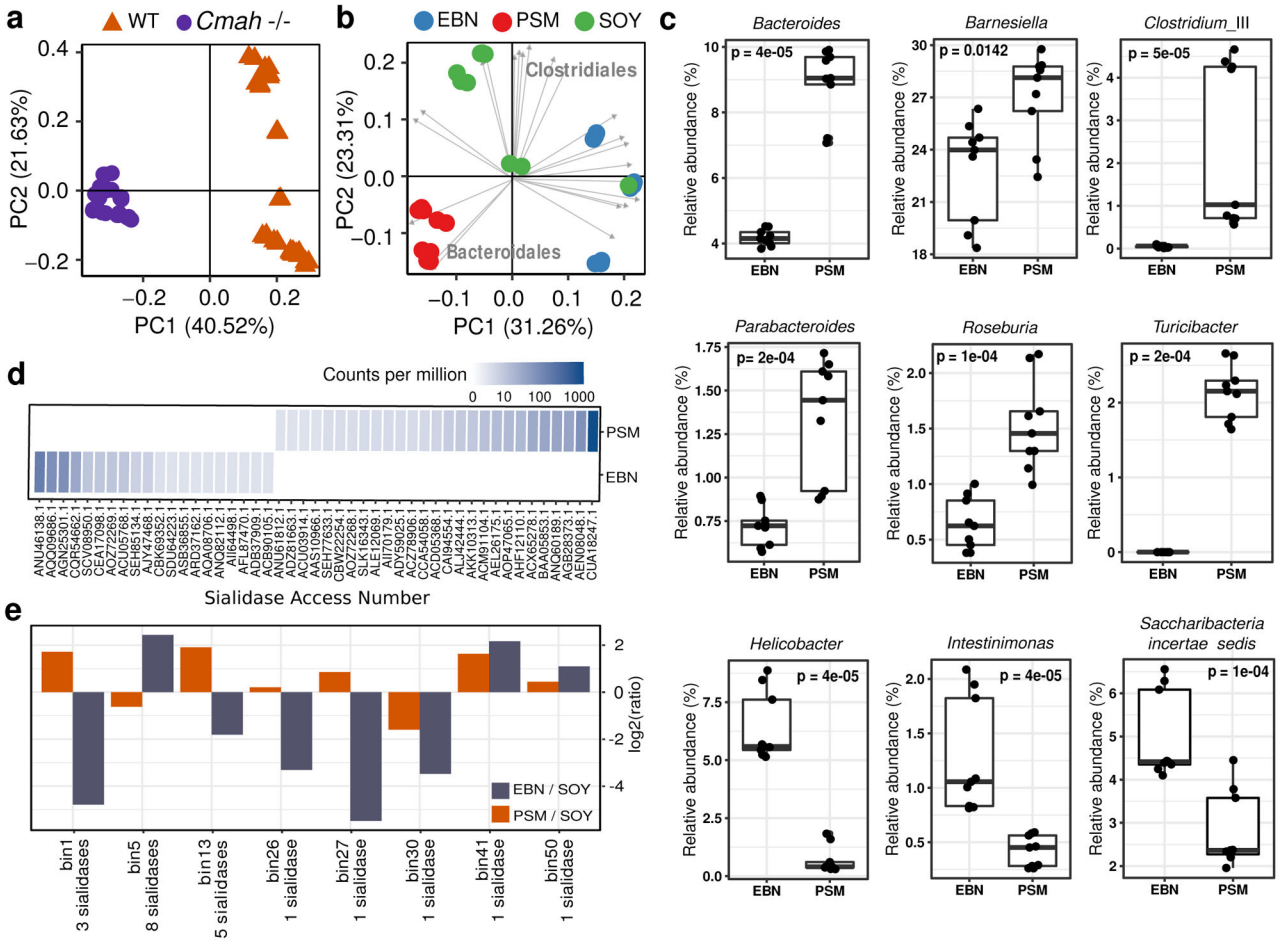

Fig. 1: Composition of gut microbial community of mice fed on Soy, PSM (porcine submaxillary mucin), or EBN (edible bird's nest) diet.

a, Beta-diversity analysis of wild-type (WT) vs *Cmah*$^{-/-}$ mouse. Pairwise Bray-Curtis dissimilarities were plotted against the first and second principal coordinates (ANOSIM R=0.979, *p-value*=0.001). WT samples include n=5 biologically independent animals and n=15 independent experiments per diet. *Cmah*$^{-/-}$ samples include n=3 biologically independent animals and n=9 independent experiments. **b**, Beta-diversity analysis of *Cmah* $^{-/-}$ mouse fed with soy, PSM, or EBN diet. Pairwise Bray-Curtis dissimilarities were plotted against the first and second principal coordinates (ANOSIM R=0.831, *p-value*=0.001). Gray arrows represent the significant vector fitting with each of the PcoA (Principal Coordinates Analysis) ordinations. The most representative taxa are indicated on the plot. **c**, Differentially abundant bacterial genera on PSM and EBN diet. Boxes and whiskers indicate quartiles and middle maker indicate the median (n=9). *p-values* were determined using the two-sided Wilcoxon rank sum test with Holm correction for multiple hypotheses. **d**, Sialidases diet-dependent using CAZYme database. **e**, Relative abundance ratio (PSM/SOY and EBN/SOY) of the bins with annotated sialidases.

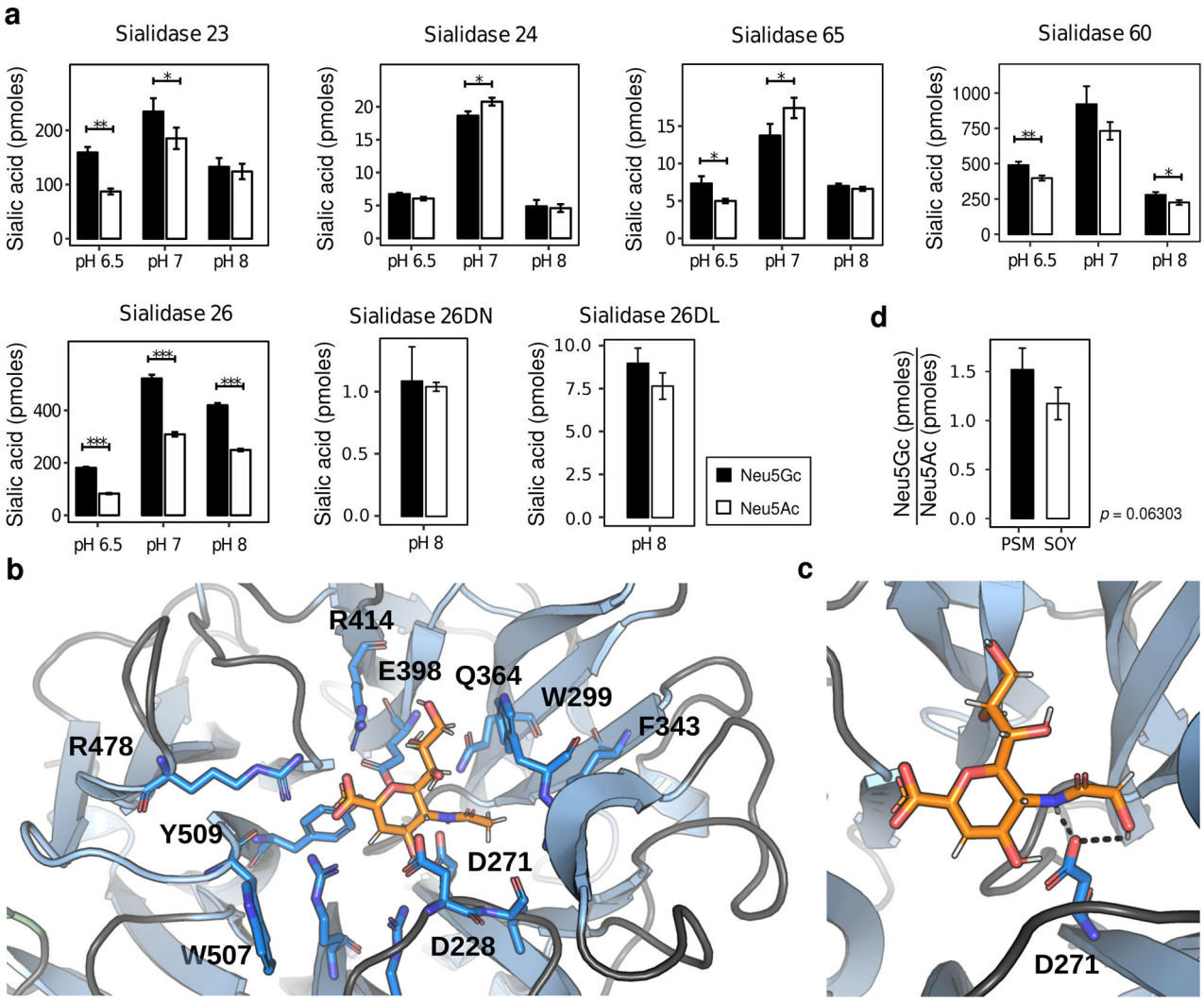

Fig. 2: Characterization of sialidases preference for Neu5Ac or Neu5Gc-containing substrates.

a, Sialidase activity assay using 0.5μg of each sialidase in three different pHs (6.5, 7.0, and 8.0). A mix of WT and *Cmah*$^{-/-}$ mouse serum containing similar amount of bound-Neu5Ac and -Neu5Gc was used as substrate. Data are represented as triplicates (n=3 biologically independent experiments). **b,** Crystal structures of sialidase26 co-crystallized with DANA-Ac. **c,** Crystal structures of sialidase26 co-crystallized with DANA-Gc. Interacting side chains are represented as sticks (blue), as well as the ligand (orange). (**b** and **c**), The protein residues interacting with the sialic acid and the residue position (such as R478 and D271) are indicated in black. **d**, Overall ratio sialidase activity in clarified fecal samples from mice fed with PSM (n=4 biologically independent experiments) or Sialic acid-free soy (n=3 biologically independent experiments). Sialidase activity assay was performed using 2μg of total protein and a mix of WT and *Cmah*$^{-/-}$ mouse serum containing similar amount of bound-Neu5Ac and -Neu5Gc. Sialidase activity is presented as a ratio of Neu5Gc (pmoles) and Neu5Ac (pmoles). **a, b, and d**, Statistical significance were determined by Student's 2-tailed t test. The significance levels are indicated as follow: $p\text{-value} \leq 0.05$ (*), $p\text{-value} \leq 0.01$ (**), and $p\text{-value} \leq 0.001$ (***). Bars represent geometric mean ± s.e.m.

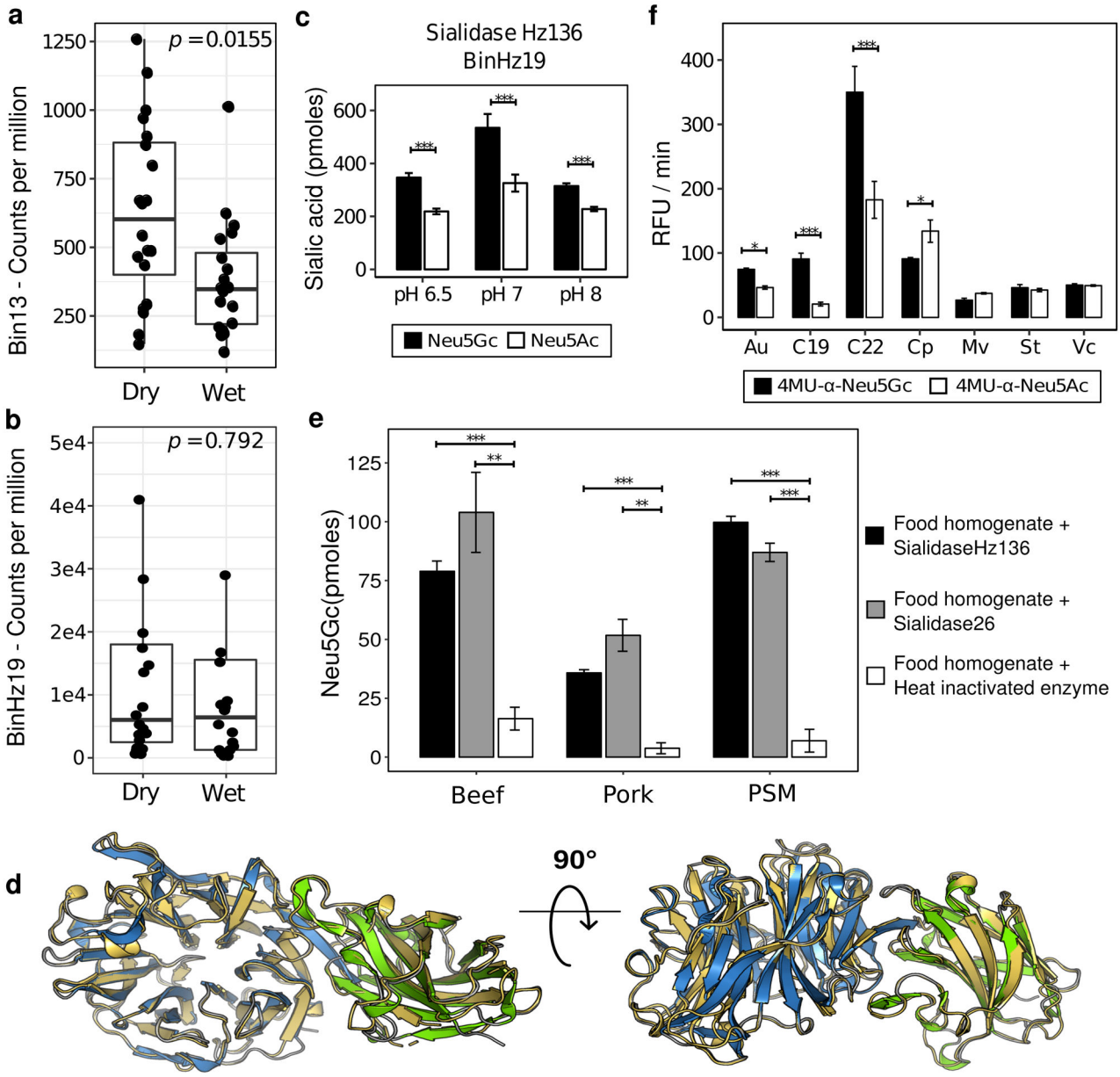

Fig. 3: Screening for Neu5Gc-preferring sialidases in human and environmental samples.

a, Relative abundance of bin13 in samples from Hadza hunther-gatherers population (n=20 biologically independent samples per season) **b,** Relative abundance of binHz19 in samples from Hadza hunther-gatherers population (n=20 biologically independent samples per season). **c,** Activity of sialidaseHz136 retrieved from Hadza shotgun metagenome in three different pHs (6.5, 7.0, and 8.0) using 0.5μg of each enzyme. Data are represented as triplicates (n=3 biologically independent experiments) and the statistical analysis were performed by Student's 2-tailed t test. **d,** Crystal structure of sialidaseHz136 (yellow) aligned to the catalytic site (blue) and carbohydrate-binding motif (green) of sialidase26. **e,** Sialidase26 and sialidaseHz136 activity on beef (New York steak), pork (pork sausage), and PSM chow. Data are represented as triplicates (n=3 biologically independent experiments). **f**,

Comparison of composting sialidases activity using fluorogenic 4MU-α-Neu5Gc and 4MU-α-Neu5Ac substrates. RFU (Relative Fluorescence Units). The experiments were performed in duplicates (n=2 biologically independent experiments). Au – *Arthrobacter ureafaciens*, C19 – composting sialidase C19, C22 - composting sialidase C22, Cp – *Clostridium perfringens*, Mv – *Micromonospora viridifaciens*, St – *Salmonella typhimurium*, and Vc – *Vibrio cholerae*. **a and b**, Statistical significances were determined using the non-parametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypotheses. Boxes and whiskers indicate quartiles and middle maker indicate the median (n=20). **c, d, and f,** Statistical significance were determined by Student's 2-tailed t test. Bars represent geometric mean ± s.e.m. The significance levels are indicated as follow: $p\text{-value} \leq 0.05$ (*), $p\text{-value} \leq 0.01$ (**), and $p\text{-value} \leq 0.001$ (***).